AD P002875

DOING C^2 EXPERIMENTS USING WAR GAMES

Dr. Joel S. Lawson, Jr.
Technical Director, C^3I Systems & Technology Directorate (ELEX 06T)

Naval Electronic Systems Command
Washington, DC 20363

INTRODUCTION

In the last few years considerable progress has been made in the development of an analytic theory of military Command Control (C2), both as a process and as a large-scale system. There are now models of C2 organizations which permit the examination of the effects of various changes in a C2 system and which can predict some of the behavior of such a system in a gross sense. And, due to the increased attention being given the field, with the attendant increase in papers, workshops, etcetera, there is slowly developing a common vocabulary for use in the emerging C2 Theory.

What is still lacking, however, is a body of experimental data which can be used as a "touchstone" to guide further theoretical developments, and against which theoretical predictions can be tested.

This paper reports the results of a very rudimentary experiment which was conducted at the Naval Postgraduate School at Monterey, California, during the 1983 Winter Quarter to test two specific hypotheses. As is often the case, it was found necessary to modify or restate the hypotheses during the conduct of the experiment in order to accommodate certain "real world" constraints. The results, however, are both interesting in their own right and reassuring for the prospect of being able to do further experiments in the C2 arena.

Motivation for the Experiments

The practical motivation for this experiment lay in a desire to study the "decision making" process in a C2 process. "Decisions" are one of the two major products of a C2 system. (The other major C2 system product may be considered to be "plans" which are statements of actions to be taken by various parts of the system. They are generated either to implement a previous decision or to provide for possible future decisions. In many cases they have the nature of "pre-determined decisions.") It was this motivation which led to the approach and the experimental procedure adopted.

In keeping with a "hard science" approach to the development of a C2 Theory, we wanted to find "observables" which could be observed by people who were not part of the system and which could also be observed by other investigators, in other experiments.

It should be noted that, by the nature of the C2 business, which involves both people and complex scenarios. it is nearly impossible to "replicate" an experiment. People, unfortunately, learn, and their behavior on a second trial will be different from what it was on the first. Also, different "command teams" or battle staffs will have different approaches or "styles" in dealing with the same basic problem. Furthermore, in "free play" exercises, which are the only ones in which real decision making takes place, the dynamics of the entities in the scenario very seldom repeat. In particular, successive plays of the same scenario generally do not lead to the same physical configuration of the battlefield after even a few minutes of play. Therefore, "replication," in the sense it is used in physics or chemistry, is a practical impossibility. Rather, we must search for observables which are in some way connected to the decision-making process and which can be observed in numerous settings, but which are independent of the details of the scenario and the particular participants.

Pursuing this line of reasoning, it was decided that an interesting characteristic quantity might be the time interval between decisions, independent of what the decisions were. A histogram of this interval against the number of cases in which it was observed should give one some idea about the tempo of activity in a command center.

But this leads to the question of defining "a decision" and how it is to be observed. At first it was felt that decisions could be grouped into broad classes, such as tactical, strategic, or logistics, and observed or logged by listening to the conversations between members of the battle staff. In practice, this turned out to be virtually impossible, and a simpler but more exact definition of a decision was adopted.

Specifically, decisions were defined as space-time events made manifest by either utterances or a series of keystrokes on a computer terminal which request or transmit information or directions. This definition satisfies the requirement that the decision be observable, i.e., it is given physical reality. (My decision to have steak for dinner tomorrow has no reality until I make it known to the outside world by telling someone else.) This choice also takes us down to the "operating level" in the decision-making process. A decision to ask for information, or to send a message, is now treated equally with a decision to engage the enemy. (Perhaps a better name for "decisions" defined in this way would be "transactions," but for now we shall continue to call them decisions.) This definition of decision led directly to a revised version of the original first hypothesis:

> A Commander or Battle Staff produces decisions which are evidenced by utterances or keystrokes and which can be observed by external observers. Furthermore, different external observers will agree on the "class" to which the observed decision belongs if the classes are chosen in a reasonably general manner.

The notion of a "class" of decisions will be described further below, when we discuss the experimental procedure. But before going on to that, it will be recalled that there was a second hypothesis to be tested in this experiment. The original form of this hypothesis was stated as follows:

> The histogram of time intervals between decisions should give an indication of the state of training or competency of the Commander or Battle Staff.

It seemed reasonable to assume that a relatively untrained team would take almost random amounts of time to arrive at each of a series of decisions, leading to a nearly flat histogram. At the same time, a more experienced group would have better defined work habits and evidence a more peaked distribution. In fact, one might expect a bimodal distribution, with simple or routine decisions being made quickly and effortlessly, while more complex procedures took longer. Alternatively, a complex problem might result in a long time delay, followed by a flurry of activity with short apparent inter-transaction times.

Unfortunately the data collected in this experiment do not allow for a simple determination of which decisions may have resulted from some previous one. Nor is there any direct way to associate one decision with another in an input-output sense. Investigation of these matters will have to await an improved experimental procedure.

The Experimental Setting

An opportunity to test these hypotheses arose during the 1983 Winter Quarter at the Naval Postgraduate School, Monterey, California.

A series of computer-based war games were to be run, using the Warfare Environment Simulator (WES) at the Naval Ocean Systems Center (NOSC), San Diego, California, to which the Postgraduate School has a remote access capability. The purpose of these games was to give the students in the Operations Research curriculum an opportunity to get "hands-on" experience with the operation of a war game. Thus, the subject "Battle Staffs" were certainly inexperienced and, in fact, being made up of members of all four services, many of the players were unfamiliar with Navy terminology and procedures.

The WES facility at the Postgraduate School at that time consisted of three sets of displays, one each for the Orange and Blue teams and one for the umpire or game coordinator. Each set of displays had a large graphic display of the theater of operations in the center with a smaller alphanumeric display and keyboard on each side of it. The scale of the geographic display could be changed to provide a "zoom" capability and appropriate symbology depicted the location of objects. One of the alphanumeric displays was generally used only as a status board, while the other was used to give instructions to the machine to effect control of that side's forces.

The display systems in the Blue and Orange command posts showed only those targets which had been detected by their own sensor systems, while the umpire's position had access to the data as seen by each side as well as to ground truth.

The class running the war game was split into three groups (A, B, and C) of two (Orange and Blue) three-man teams. Each was a familiarization session in which they learned how to operate the keyboard and make the machine do some elementary things. In the second session they actually did play a war game, deploy their forces, and get to the point of engaging the enemy. The final session was run "for the record" and was based on a scenario and initial force disposition very similar to the ones with which they had practiced. By this time the Battle Staffs seemed to be reasonably comfortable with the system and could concentrate most of their effort on the battle rather than the game mechanics.

To carry out the experiment and test these hypotheses, observers were recruited from other students enrolled in the Command Control and Communications curriculum at the Postgraduate School, and arrangements made to have them present during the running of the games.

Only limited observations were made during the first series of runs, to test the methodology. The data reported here were all taken during the second and third series of runs, when the performance of the Battle Teams had become reasonably stable.

Experimental Procedure

For the experiment, each observer was equipped with a clipboard of data sheets which had a synchronized clock mounted on it. They were instructed to record to the nearest second when they thought they observed a "decision" and the "class" of that decision. A sample of the data sheet is reproduced in Figure 1, which also shows the classes of decisions used in this experiment. These classes were intended to be generic in nature and encompass nearly all the situations which were expected to be observed.

The observer was also asked, if he had time, to enter any amplifying information in the "Notes" column. The columns for "Game Time" and "Enter Time" were provided to allow correlation to game time (which can be faster or slower than clock time) and to note the time when an instruction which had been observed passing from one team member to another, was actually entered on the keyboard. The latter turned out to be an impossible task as the keyboards were in such constant use as to preclude connecting any particular set of keystrokes with any specific utterance. (In future experiments it may be possible to achieve some of these measurements by installing appropriate "hooks" in the game software.)

After the experimental session, the observations were transcribed into a computer program which computed the desired time intervals and could generate various histograms of the results. Figure 2 is a sample of a data file generated by this program.

As shown in the Figure, it was eventually decided to classify the decisions by "type" as well as "class." This was partly due to the confusion between different observers as to which class a particular decision should be placed in, and partly because of the limited number of observations of members of a particular class. By introducing a definition of "types" of decisions as either "information decisions" or "action decisions," the sample size was effectively increased and the differences between observers was reduced to a negligible amount. Examples of this distinction are decisions which deal with "information," such as requesting status or identity, and those decisions which deal with "action," such as deploying a force or changing the EMCON condition.

With these additions to the data definitions, the time intervals are defined as follows:

T0 = real world clock time
T1 = time since last decision, independent of its class
T2 = change in T1 since last T1 event
T3 = time since last decision of same "type"
T4 = change in T3 since last T3 event
T5 = time since last decision of same "class"
T6 = change in T5 since last T5 event

By generating such data files on disk, each representing the observations of one observer (using the observer's name as file name) during a game he observed (using the team name and game number as a file extension), the data was set up to allow comparison of different observers as well as different teams and games.

The results of this analysis are reported in the next section.

Experimental Results

Before examining the specific experimental results, a word about the choice of "classes of transactions" is in order because they turned out to influence the apparent results. In particular, the distinction between ordering the deployment of one or two units or platforms and ordering the deployment (e.g., a course change) of the whole force was interpreted differently by different observers. And the same was true of requests for the status of individual units or of a major portion of the force. Much of this confusion could have been overcome if there had been more time to test the methodology and have the data takers agree on some conventions. As it was, there was a minimum amount of coordination between the observers, and each one made up his own rules as to how he would classify the events he observed.

The basic classes were originally chosen because it was assumed that they were sufficiently general to encompass all the transactions which were likely to take place, while also being representative of distinctly different sorts of activity which one would expect to see in a command center.

In practice, it appears that these were reasonable assumptions, but that the definition of the "classes of transactions" can probably be improved, and in fact should probably be tailored to match the nature of the scenario and/or level of command which is being examined. Figures 3 and 4 present a typical comparison of the time between transactions, independent of the type or class of transaction, as recorded by two different observers watching the same battle staff. These Figures are histograms of the time intervals between observed decisions, grouped into 18-second wide bins. (The units of "width" are seconds.) For ease of comparison, the actual number of observations has been normalized to 100 so these represent percentage distributions.

It is immediately obvious that the observers were apparently observing the same sort of behavior and recording a similar series of events. A detailed comparison of their event logs bears this out. In fact, in all cases where there were two or more observers watching the same battle staff, there was remarkable agreement in what they logged as "decisions" or transactions.

The two major sources of nonconformity of the logs were the differences of interpretation mentioned above and a very simple but important physical limitation. Because the battle staff were sitting side-by-side in front of the displays, the observers tended to sit on either side of them so they could both see the displays and hear what the group was talking about. This led to one observer being able to hear remarks about what was on the status board better than the other, while the second was more alert to directions to the keyboard operator.

Nonetheless, as a general rule the overlap of the logs of pairs of observers was about 60 percent, even without any serious attempt to agree on just how the classification scheme would be applied. At the grosser level of whether it was an "action" transaction or one dealing with information, the agreement was nearly 80 percent. The "disagreements" are nearly all accounted for by a missing entry on one of the logs. That is, one observer logged an event which the other did not. In most cases this can be explained as a result of the placement of the observer relative to the battle staff, as mentioned above.

Moreover, data taken by the same observer watching different Battle Staffs produces plots very similar to those shown in Figures 3 and 4. So the distribution of time intervals between decisions does seem to be a characteristic quantity, at least in this setting.

An interesting change is observed, however, when the time between successive action or information transactions is plotted, as in Figures 5 and 6. The action decisions still show the Rayleigh-like distribution, while the information decisions are more uniformly distributed. No explanation for this is offered, but an interesting comparison with some other data will be made below. (Also it was noted that some observers seemed to be more sensitive to action decisions than to information decisions, perhaps reflecting their personality more than their location in the command space.)

Based on these results, it seems that the first of the hypotheses to be tested is confirmed by the data which was taken during this experiment.

As far as the second hypothesis is concerned, we can only report that there is no significant difference between the data obtained in the second and third sets of war games. If there was any "learning effect," it was masked by other attributes of the game.

A final interesting observation is that in essentially all cases the observers logged about two-thirds of the transactions as action decisions and only one-third as information decisions. This may be partly due to the nature of the WES facility, in which the Battle Staffs have to issue a rather large number of maneuvering instructions to their (computer-simulated) forces. This has the effect of putting the Battle Staff in a much more "tactical" role than a "command" role. Also, with only one command center on each side, there were no superiors or subordinates to answer questions for or receive queries from. A third cause may have been the presence of electronic status boards which provided information without the verbal clues which permitted the transaction to be recorded.

Comparison With a Manual War Game

Coincident with the conduct of this experiment, the "historical logs" of a game run at the war-gaming facility at the Naval War College at Newport, Rhode Island, during the preceding fall became available.

This game also involved some inexperienced players but these were supported by trained and experienced Naval officers, and both sides had had considerable time to develop their plans before the game commenced. In addition to the game being basically a manual one without automated status boards, etc., there were several levels of command represented on each side. Thus it was quite a different environment and perhaps more closely simulated the conditions which one might expect in a typical command node in a Navy C2 system.

In order to obtain some comparative data, the logged events were divided into the classes shown in Figure 7. These classes were chosen as being more representative of the log entries, although they still retain the distinction between action and information decisions. (Although these classes are also divided into "input" and "output" groups, no analysis of this dimension has yet been undertaken.)

Using these data, which are based on the editing and compilation of many observers' logs into one by the game historian, similar analyses were carried out. Two typical plots are shown in Figures 8 and 9, which depict one six-hour period in the game at two levels of command, CTF 70.1, the Battle Force Commander, and 70.X, one of his four immediate subordinates. It should be noted that the basic time unit of "width" used in these Figures is one minute rather than the second used for the WES data.

While the most striking thing about these data is the similarity of the plots to those obtained with the data taken during the WES war games, there are several other interesting points.

First, the activity seems to be peaked at the level of the Battle Force commander. While not displayed, the plots of activity at the level of SEVENTH Fleet Commander show only about one-sixth the number of decisions per unit time that are evident at the Force Commander level. And the Force Commander's subordinate is operating at about half his tempo. However, there are four subordinates, so the total activity is higher at the lower level, as one would expect it to be.

Second, the effective mean time between decisions seems to be about two minutes, rather than the 15 to 20 seconds seen in the WES game. This, of course, may be an artifact of the coarser time increments used in the Newport logs, as well as the less tactical nature of the decisions being made.

And thirdly, the statistics on types of transactions were exactly reversed from that seen at WES. Two-thirds of the decisions involved information and only one-third or less involved action. Whether this was due to the lack of computer aids or to a more realistic simulation of a command situation is not clear. It is hoped that additional light can be shed on this subject by obtaining some more data at the War College, which has had a major computer facility installed to support its war gaming functions since these data were taken.

Conclusions

First, it seems evident that one can define space-time events (which we shall call decisions or transactions) which deal with either actions or information and which are made observable through utterances or keystrokes made by the Battle Staff, on which different observers will agree. That is, different observers will report observing the same type of event at the same time. And it seems likely that these space-time events could be further subdivided with more careful training and discussion among the observers.

Second, the gross inter-transaction time has a distribution not unlike a Rayleigh distribution, and which seems to be a common behavior even in quite different situtations.

Third, the inter-transaction time between transactions of the same type may differ between different types.

Fourth, the time between transactions seems to be a minimum at the level of the highest "on-scene" Commander. This is interpreted as meaning that his "operating tempo" is forced to be high enough to encompass the total activity below him, which is divided among several subordinates.

And finally, the qualitative differences between the Newport data and that taken on WES may indicate that introducing computer assistance into a command center will change the nature of the "transactions" which one sees taking place. If true, this may have serious implications for the design and organization of future command control systems.

```
==========================================================================
Group:       Side:       Date:        Observer:                   Page:
--------------------------------------------------------------------------
Decision Classes:
  ES = request EMCON status        CE = order a change in EMCON
  FS = request Force status        DF = deploy the Force
  US = request status of a unit    DU = deploy a unit
  RI = request for an identity     EE = engage enemy (or target)
  SD = seek other data             AO = all other decisions
==========================================================================
Decision | Decision |                  Notes                | Entry | Game
  Time   |   Type   |                                       | Time  | Time
==========================================================================
```

Figure 1.
Data Form Used for WES Experiments.

COMMAND CENTER INTER-ACTIVITY TIMES

No.	DTG	T0	T1	T2	TYP	T3	T4	Act	T5	T6
0	151800	0	0	0	0	0	0		0	0
1	151833	33	33	33	2	0	0	AO	0	0
2	152317	317	284	251	1	0	0	FS	0	0
3	152406	366	49	-235	2	333	333	AO	333	333
4	152432	392	26	-23	1	75	75	FS	75	75
5	152508	428	36	10	1	36	-39	FS	36	-39
6	152530	450	22	-14	2	84	-249	DU	0	0
7	152612	492	42	20	1	64	28	SD	0	0
8	152735	575	83	41	1	83	19	US	0	0
9	152807	607	32	-51	1	32	-51	SD	115	115
10	152854	654	47	15	1	47	15	US	79	79
11	152937	697	43	-4	2	247	163	AO	331	-2
12	153015	735	38	-5	2	38	-209	AO	38	-293
13	153058	778	43	5	2	43	5	DF	0	0
14	153146	826	48	5	2	48	5	DF	48	48
15	153256	896	70	22	2	70	22	AO	161	123
16	153323	923	27	-43	1	269	0	FS	495	0

Figure 2.
Print-Out of Computed Time Intervals as Stored in Computer Data Files.

```
0        10        20        30        40        50
I....I....I....I....I....I....I....I....I....I....I
I******************
I********************************
I*******************
.**************
I***
I******
I***
I*
I
I***
I
I
I*
129 Observations 116 Samples  12 Bins      Width = 18
File = WREN/B2      Act.type = 1      Time int.= T 1
Game interval from: 153000 to: 170500      NORMALIZED
```

Figure 3.
Distribution of Time Intervals Seen by Observer Wren.

```
0          10         20         30         40         50
I....I....I....I....I....I....I....I....I....I....I
I****************
I****************************
I******************
I*************
I*******
I*******
I*****
I
I**
I
I*
I*
I**
105 Observations 96 Samples   12 Bins     Width = 18
File = LOMBARDO/B2 Act.type = 1      Time int.= T 1
Game interval from: 153000 to: 170500     NORMALIZED
```

Figure 4.
Distribution of Time Intervals Seen by Observer Lombardo.

```
0          10         20         30         40         50
I....I....I....I....I....I....I....I....I....I....I
I*************
I*************************
I****************
I**********
I**
I***
I*****
I***
I*
I**
I
I
I*****
129 Observations 85 Samples   12 Bins     Width = 18
File = WREN/B2     Act.type = 3      Time int.= T 3
Game interval from: 153000 to: 170500
```

Figure 5.
Time Intervals Between "Action" Decisions (Un-normalized).

```
0          10         20         30         40         50
I....I....I....I....I....I....I....I....I....I....I
I*
I***
I**
I****
I
I***
I*
I*
I**
I*
I**
I*
I**********
129 Observations 31 Samples   12 Bins     Width = 16
File = WREN/B2     Act.type = 2      Time int.= T 3
Game interval from: 153000 to: 170500
```

Figure 6.
Time Intervals Between "Information" Decisions (Un-normalized).

	Input transactions	Output transactions
Info:	RR = receive a report RQ = receive a query	SR = send a report SQ = send a query
Action:	RD = receive a directive RT = receive tasking	SD = send a directive ST = send tasking
Miscel	AO = all other decisions	JE = journal entry

Figure 7.
Decision Taxonomy Used for Newport data.

The distinction between directives and tasking is that directives leave the method of implementation to the subordinate, while a tasking includes specific detailed instructions. The JE category was used for remarks by the historian.

```
0         10        20        30        40        50
I....I....I....I....I....I....I....I....I....I....I
I********************
I****************************************************************
I*****************************************
I***********************
I************
I*****
I***
I**
I
I*
I
I*
I**
174 Observations 174 Samples  12 Bins     Width = 1
File = CTF701/P1   Act.type = 1      Time int.= T 1
```

Figure 8.
Distribution of Time Intervals Observed in
Command Center of Battle Force Commander.

```
0         10        20        30        40        50
I....I....I....I....I....I....I....I....I....I....I
I*********
I**************
I********************
I********
I*********
I*******
I*******
I***
I**
I*
I*
I**
I******
89 Observations  89 Samples   12 Bins     Width = 1
File = CTF703/P1   Act.type = 1      Time int.= T 1
```

Figure 9.
Distribution of Time Intervals Observed in the
Command Center of a Subordinate Battle Group Commander.